Determining Shoal Membership: A Comparison between Momentary and Trajectory-

based Methods

Vicenç Quera[a], Francesc S. Beltran and Ruth Dolado

Institute for Brain, Cognition and Behavior,

Adaptive Behavior and Interaction Research Group,

Department of Behavioral Science Methods, University of Barcelona,

Campus Mundet, Passeig Vall d'Hebron 171, 08035 Barcelona, Spain

[a] Corresponding author. Phone: +34 933 125 088. Fax: +34 934 021 359. Email:

vquera@ub.edu

Abstract

Miller and Gerlai proposed two methods for determining shoal membership in zebrafish (*Danio rerio*), based on momentary mean inter-individual distances and on post-hoc analysis of trajectories of nearest neighbor distances, respectively. We propose an alternative method based on momentary nearest neighbor distances and compare the three methods using simulation. In general, our method yields results more similar to Miller and Gerlai's second than first methods, but is computationally simpler.

Keywords

Coordinated collective motion is a common phenomenon in numerous animal species, consisting of many individuals adjusting their movements to achieve coherent group travel. Coordinated motion exists when: (a) the group is polarized or aligned, i.e., most individuals move with a similar heading, and change heading in a synchronized way; (b) most individuals adjust their speed to the average group speed; and (c) individuals tend to aggregate, while keeping some distance from their neighbors to avoid collision. Bird flocks and fish schools are two prominent instances of coordinated collective motion, whose functionality, evolution and underlying organizational mechanisms have been the subject of many studies in recent decades (e.g., [1, 2, 3, 4]).

Two kinds of group organization are usually described in fish: shoals and schools. According to Pitcher and Parrish [5], "groups of fish that remain together for social reasons are here termed shoals (…) synchronized and polarized swimming groups are termed schools. Schooling is therefore one of the behaviors exhibited by fish in shoals, and schools have a structure measured in polarity and synchrony" (p. 365). Miller and Gerlai [6, 7] studied the temporal dynamics of shoal organization in zebrafish (*Danio rerio*), specifically how the average distance among fish in the shoal changes and oscillates through time in the presence of food and predators. To that end, they tracked in 2D individual fishes in an experimental group of 16 using specific video-analysis software to obtain inter-individual distances by sampling one frame per second during 1-min intervals. As some fishes were frequently observed to swim away from the main group, thus inflating the average distance, they devised an objective method [7] for deciding which individuals could be considered members of the shoal, so that only distances among these fish would be taken into account when the average was calculated and, consequently, deflated. Recently, Miller and Gerlai [8] presented a new method based on the statistical distribution of nearest neighbor distances; unlike their

previous method, which toke the momentary spatial distribution of the individuals into account only, the new one was based on trajectories, and defined group-membership thresholds based on how distances evolve throughout the session. They studied how individual excursions (i.e., out of the shoal) drive the shoal dissolution, and how the location of the individuals in the shoal is related to their tendency to initiate excursions from it. To that end, they tracked in 2D individual fishes in groups of 8 by sampling frame by frame.

Using specific agent-based simulation software (MovAgent; www.ub.edu/gcai/group/Downloads.html), we simulated the emergence of schooling in groups of different sizes, and applied Miller and Gerlai's methods for determining shoal membership; our software and the model on which it is based were described previously [9]. Briefly stated, virtual agents' motions are governed by local rules determining how each agent adapts to others it perceives at each time unit by changing the ideal distances it wants to keep from them; subsequent motion results from attempting to minimize its *dissatisfaction*, defined as a function of the discrepancy between its real and ideal distances. Ideal distances between agents vary as a consequence of the outcomes of their interactions, according to a set of reward rules. Unlike Reynolds's [10] and Huth and Wissel's [11] flocking and schooling models, these rules are not directly related to the parameters that define coordinated motion at a global level (polarization, speed adjustment and aggregation), but simply specify how ideal distances change; nonetheless, when our model is simulated, coordinated motion emerges, that is, the virtual agents tend to aggregate (i.e., form shoals) and polarize much like real fish do.

We observed that, when Miller and Gerlai's earlier method [7] is applied for several successive time units, the assignment of individuals to the shoal tends to be unstable in certain cases: as individuals move, some of them may be considered

members of the shoal at a certain time unit and excluded from it at the next, while for an observer no substantial differences in membership can be distinguished (see an example in Figure 1). In such cases, individuals that are far away from the rest of the group would be alternately included and excluded from the group when analyzing distance, thus potentially altering any measures that can be obtained from the distances (e.g., oscillations in average distance). As we have observed in our simulations, this instability occurs more often in large groups than in small ones; thus, an alternative method yielding a more stable assignment of individuals to the shoal in the aforementioned cases is necessary, particularly when the number of individuals being tracked increases. Such instability might be a consequence of using both momentary mean inter-individual distances and a fixed threshold for determining group membership. In Miller and Gerlai's [8] new method, an individual is considered to be on an excursion when the distances to its nearest neighbors are greater than some thresholds for a certain period of time, the thresholds being modes and percentiles of the distributions of those distances obtained for the entire session, not for a specific moment in time. However, that method requires a post-hoc analysis of the trajectories of the individuals in order to obtain such distributions, thus preventing it to be used for determining group membership while the individuals are being tracked.

We propose a method that, like Miller and Gerlai's [8] new one, uses nearest neighbor distances; like their previous method [7], ours is based on fixed thresholds that are computed at the specific moment of time though, unlike it, thresholds are applied to raw, non averaged distances. On the one hand, thresholding individual raw versus averaged distances may reduce the instability in the number of individuals that are assigned to the shoal; on the other, using momentary thresholds is simpler than, and does not require, knowing how distances are distributed.

Miller and Gerlai's [7] method is as follows. Given a group of N fish, the mean distance from every fish to all others is computed at a certain time unit; mean distances are ranked from lowest to highest, and differences between every pair of consecutive ranked distances are computed. Starting at the fish with the lowest mean distance, subsequent individuals are considered to be members of the shoal up to that individual for which the difference is greater than a certain momentary threshold criterion, the square root of the current global mean distance among all the fish, $\sqrt{D}$.

Miller and Gerlai's [8] newer method is based on nearest neighbor distances (NND) instead: (a) in each unit of time, NNDs are computed for every individual; (b) the distribution of the NNDs is obtained by pooling the data over all the individuals throughout the session, and its mode M is computed; (c) the time series of NND for each individual is analyzed, and the segments starting when NND $\geq$ M and ending when NND < M are selected; (d) for every selected segment, the maximum NND reached within it is computed (MaxD); (e) the distribution of those maxima is obtained by pooling them over all segments and individuals, and its 95th percentile (P) is computed; (f) finally, all segments for which MaxD $\geq$ P are classified as excursions from the shoal. Generalizing for the k-th nearest neighbor ($k = 1 \ldots N/2$), the procedure can be repeated for NND_k, by computing the differences $dNND_k = NND_k - NND_{k-1}$ (with $NND_0 = 0$) in each time unit and, subsequently their distribution and mode. That way, the segments during which each individual is away from the shoal on a "type k excursion" can be obtained. Individuals can be simultaneously on more than one type of excursion; during a type 1 excursion an individual is alone and away from the shoal, whereas during a type 2, but not a type 1, excursion it is close to another individual and both are away from the shoal. Those individuals that are not in any kind of excursion are members of a shoal at that moment; however, when several distinct shoals exist, individuals belonging

to one of them are likely to be on excursions of higher order k with respect to those belonging to other shoals, and not to be on excursions of lower order k (with respect to individuals in their own shoal). Therefore, while individuals not being on lower order k excursions are members of a shoal, not all of them necessarily belong to the same shoal, as there may be several.

The method we propose is as follows: (a) in each unit of time, k-th nearest neighbor distances are computed for every individual, NND_k ($k = 1, …, N−1$); (b) differences between every pair of consecutive distances are computed; (c) an individual is away from the shoal if, starting at $k = 1$, the first difference $dNND_k$ that is greater than a criterion threshold ($\sqrt{D}$, as in [7]) is for a k that is less than a criterion percentage F of the group size N ($F = 20\%$). In our simulations we have observed that the combined effect of those two criteria tends to segment the group fairly well, identifying which individuals are outside, or away from, the shoal (as determined by eye), in a way analogous to a segmentation based on a bimodal distribution of NNDs, as proposed by Aoki [12].

We ran several series of simulations with MovAgent in 2D, for $N = 20$, 40 and 60 virtual agents, with five replications each, and 3,000 time units per simulation. The agents were initially scattered at random in a square virtual pool (a lattice of size 120×120 cells; maximum possible distance between agents was thus 169.7 cells); coordinated motion usually emerged after hundreds of time units, and from that point on, the simulation results were inspected by applying the three methods, which are termed M1 (Miller and Gerlai's [7]), M2 (our proposed method), and M3 (Miller and Gerlai's [8]). For each replication, six time series were obtained, with the number of individuals belonging to the shoal and with the global mean distance among them, varying throughout the simulation according to each method.

For M3, four different shoal sizes were obtained, upon discarding individuals on type 1 excursions (M3.1), individuals on types 1 or 2 excursions or both (M3.2), individuals on types 1, 2 or 3 excursions or on any combination of these (M3.3), and individuals on types 1, 2, 3 or 4 excursions or on any combination of these (M3.4); stopping at type 4 excursions is obviously arbitrary, and higher order k excursions could be taken into account. It must be stressed that defining shoal size that way does not necessarily imply that the individuals that are accounted for actually belong to a single shoal. Method M3, the way it is applied here, is not able to identify which individuals belong to each shoal. Neither is method M2, as it is based on nearest neighbor distances too. For these methods, shoal size is then the number of individuals belonging to any shoal, be it one or several. Only under the assumption of a single shoal those methods can yield a realistic shoal size.

Average shoal sizes according to each method are shown in Figure 2a, which summarizes the five replications for each group size N; shoal sizes were averaged over 10,000 time units in total, that is, over the five replications, after discarding the first 1,000 time units in each, during which coordinated motion was unlikely to have emerged yet. Method M1 is the least conservative one, yielding bigger shoal sizes than the other two, except in the case of $N = 20$, for which there is no difference between M1 and M3.1 ($z = -.80$, $p = .21$). Method M2 is the most conservative one when $N = 20$. As N increases, M3 provides more conservative shoal sizes and, as can be expected, shoal size decreases when the number of excursion types accounted for increase.

Figure 2b shows the change in shoal size over the 10,000 time units, as a function of method and group size N. The absolute value of the difference in shoal sizes between consecutive time units was computed, these values were added together, and

the sum was divided by the number of nonzero differences (i.e., effective changes). The average change in shoal size is generally much greater when M1 is applied than when M2 and M3 are applied, for all three group sizes. Statistically significant differences exist among M1, M2 and M3, and among several of the M3 sub-methods for $N = 20$ ($p < .0001$ in all cases), but not among the M3 sub-methods themselves for $N = 40$ and 60; method M3 yields very gradual changes in shoal size in comparison with M1 and M2. On the other hand, increasing the group size does not cause a remarkable effect in the average change in shoal size, for any method.

In order to know whether the mean distance is deflated when the methods are applied, we computed the mean inter-individual distance both for the group as a whole (GR) and for the individuals belonging to the shoal upon applying each method. Figure 2c shows those mean inter-individual distances (measured in lattice cells) averaged over the 10,000 time units, for the three group sizes. Applying any of the methods makes the within shoal mean inter-individual distances decrease significantly with respect to GR ($p < .0001$ in all cases); for $N = 20$, M3.4 is the method that most deflates the mean distance (significant difference with respect to M2, $z = -2.70$, $p = .003$); however, when only type 1 and 2 excursions are considered, M3 does not tend to deflate the mean distance as much as M1 and M2; for $N = 40$, the methods that most deflate the mean distance are M1, M2, and M3.4, with no significant differences among them; for $N = 60$, M2 is ($p < .0001$ with respect to the other methods). The M3 sub-methods tend to have comparable effects on deflating the mean distance as group size increases. Again, only under the assumption of a single shoal methods M2 and M3 can yield a realistic within shoal mean inter-individual distance; when several shoals actually exist, between shoal distances are likely to inflate it.

In conclusion, when M1 and M2 are compared, the latter is more conservative (it assigns less individuals to the shoal), tends to yield changes in shoal size that are more gradual, and deflates the within shoal mean inter-individual distance more (except for a medium sized group, where both methods have comparable effects). When M2 and M3 are compared, the latter is somewhat more conservative (except for a small sized group) and tends to yield changes in shoal size that are more gradual, but to deflate the mean distance less (unless up to type 4 excursions are taken into account). Comparing M1 and M3, the latter is more conservative, tends to yield changes in shoal size that are more gradual, deflates the mean distance less when group size is small (unless up to type 3 and 4 excursions are taken into account) and deflates it more when group size is medium or large.

Although, in general, M2 yields results that are more similar to M3 than to M1, it is computationally simpler than M3, our results show that neither of those methods is universally applicable for determining shoal membership, as their effectiveness depends both on group size and on the desired behavioral measure (shoal size; within shoal mean inter-individual distance) and, importantly, on the assumption that there is a single shoal plus several off-shoal individuals, which is not always necessarily the case. As the group size increases, it is likely that more than one distinct shoal appears (as determined by eye), in which case it is unclear whether the shoal size and the within shoal mean inter-individual distance provided by the methods are realistic, especially when M2 and M3 are applied; in that case, M1 is probably preferable. Nonetheless, when more than one distinct shoal actually exists, none of those methods are useful for assigning the individuals to those shoals; for example, see section 1 in Figure 1, which shows how

M1 simply discards several individuals from the shoal (a group of white dots at the left of the image), which an external observer would probably see as a second shoal.

A more refined method would be necessary to determine shoal membership in that case, that is, a clustering algorithm tailored to the analysis of spatial distribution among fish. The *affinity propagation* algorithm [13] can be a good candidate; unlike traditional clustering algorithms that have been proposed for the analysis of animal groups (e.g., the *k*-means algorithm, [14]), it does not require that the number of aggregates be specified in advance. Starting from measures of similarity among individuals (previously converted from the distances among them by means of an inverse function), the algorithm would identify how many distinct aggregates exist and assign each individual to one of them efficiently. In any case, in order to know which method for determining shoal membership is preferable, it would be desirable to compare their outcomes with a control condition obtained from estimates by observers. Both this and the development of clustering algorithms for detecting more than one shoal should be future goals in that field.

References

[1] Ballerini M, Cabibbo N, Candelier R, Cavagna A, Cisbani E, Giardina I, et al. Empirical investigation of starling flocks: A benchmark study in collective animal behaviour. Anim Behav 2008;76: 201-215.

[2] Couzin ID, James R, Mawdsley D, Croft DP, Krause, J. Social organization and information transfer in schooling fishes. In: Brown C, Laland K, Krause J, editors. Fish cognition and behavior, Oxford, UK: Blackwell; 2006, p. 166-185.

[3] Nagy M, Ákos Z, Biro D, Vicsek T. Hierarchical group dynamics in pigeon flocks. Nature 2010;464:890-894.

[4] Radakov DV. Schooling in the ecology of fish. New York: Wiley; 1973.

[5] Pitcher TJ, Parrish JK. Functions of shoaling behavior in teleosts. In: Pitcher TJ, editor. Behaviour of teleost fishes. 2nd ed. London: Chapman and Hall; 1993, p. 363-440.

[6] Miller NY, Gerlai R. Quantification of shoaling behaviour in zebrafish (*Danio rerio*). Behav Brain Res 2007;184:157-166.

[7] Miller NY, Gerlai R. Oscillations in shoal cohesion in zebrafish (*Danio rerio*). Behav Brain Res 2008;193:148-151.

[8] Miller NY, Gerlai R. Redefining membership in animal groups. Behav Res Methods in press; http://www.springerlink.com/content/f5h351046876543q/

[9] Quera V, Beltran FS, Dolado R. Flocking behaviour: Agent-based simulation and hierarchical leadership. JASSS 2010;13(2):[about 19 p.]. Available from: http://jasss.soc.surrey.ac.uk/13/2/8.html

[10] Reynolds CW. Flocks, herds and schools: A distributed behavioral model. Computer Graphics 1987;21(4):25-34.

[11] Huth A, Wissel C. The simulation of fish schools in comparison with experimental data. Ecol Model 1994;75/66:135-145.

[12] Aoki I. An analysis of the schooling behavior of fish: Internal organization and communication process. Bulletin of the Ocean Research Institute, University of Tokyo 1980;12:1-65.

[13] Frey BJ, Dueck D. Clustering by passing messages between data points. Science 2007;315:972-976.

[14] Haddadi H, King AJ, Wills AP, Fay D, Lowe J, Morton AJ, Hailes S, Wilson AW. Determining association networks in social animals: choosing spatial–temporal criteria and sampling rates Behav Ecol Sociobiol in press.

Figure captions

Figure 1

Plots of individual positions in a group of 60 at six consecutive time units, from a MovAgent simulation. Each dot represents one virtual fish; white dots represent fish that were not included as members of the shoal according to Miller and Gerlai's method based on momentary mean inter-individual distances. Note that while several individuals are excluded from the shoal at time units 1, 3, 4 and 6, they are included in it at times 2 and 5, which can be considered misassignments.

Figure 2

Comparison of methods M1, M2 and M3 for group sizes $N = 20$, 40 and 60, using data averaged over five replications with 2,000 time units each. (a) Shoal sizes according to each method; (b) changes in shoal size according to each method, sample sizes being the number of effective changes in each case, which range from 583 changes (for $N = 20$, M3.1) to 4,625 changes (for $N = 60$, M2); (c) mean inter-individual distances for the group as a whole (GR) and for the individuals belonging to the shoal upon applying each method; distance is measured in lattice cells. Error bars represent $\pm SEM$s. See text for details about M3 sub-methods.
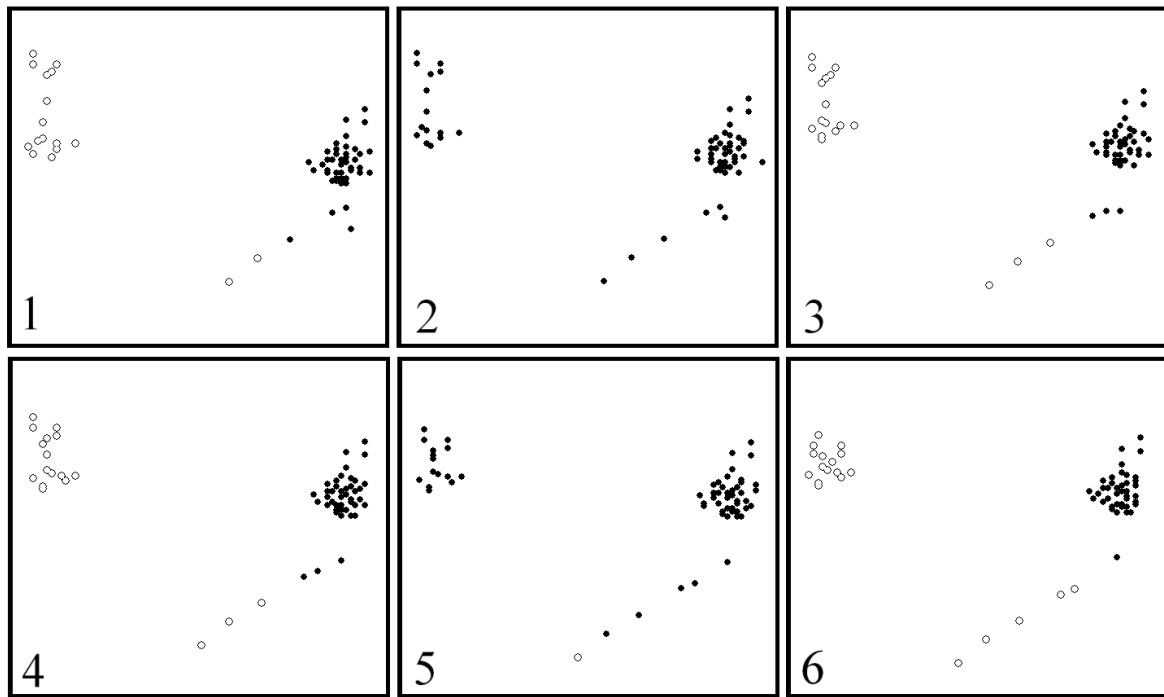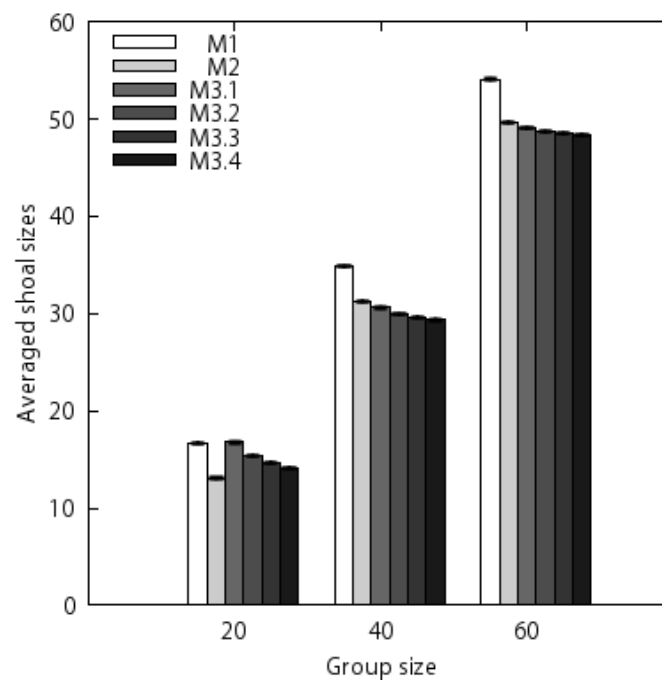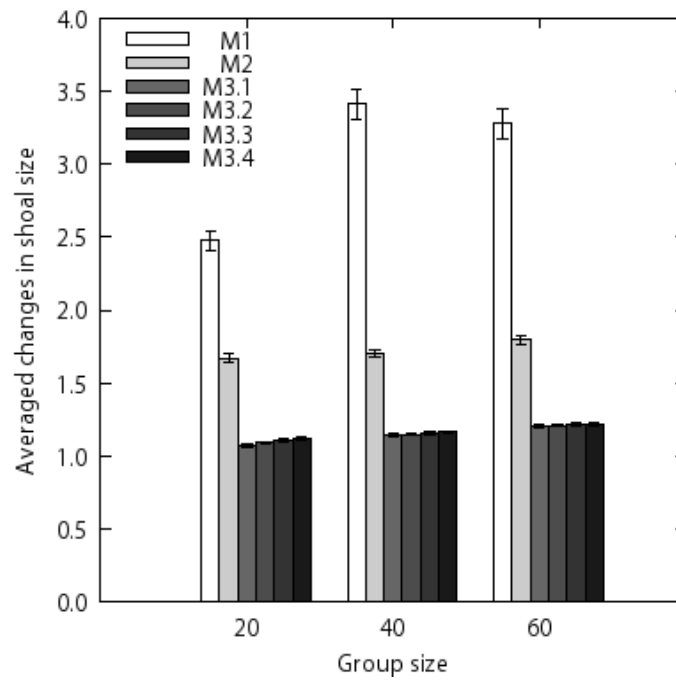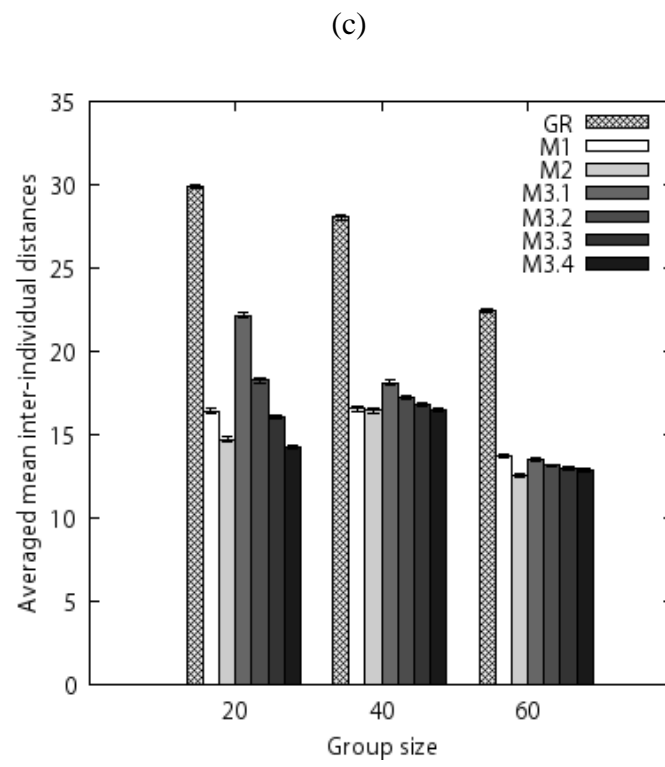
Figure 1

(a)

(b)

(c)

Figure 2